SIAM J. DISCRETE MATH.
Vol. 12, No. 4, pp. 500–523

GRAPH PARTITIONING AND CONTINUOUS QUADRATIC PROGRAMMING*

WILLIAM W. HAGER† AND YAROSLAV KRYLYUK†

Abstract. A continuous quadratic programming formulation is given for min-cut graph partitioning problems. In these problems, we partition the vertices of a graph into a collection of disjoint sets satisfying specified size constraints, while minimizing the sum of weights of edges connecting vertices in different sets. An optimal solution is related to an eigenvector (Fiedler vector) corresponding to the second smallest eigenvalue of the graph's Laplacian. Necessary and sufficient conditions characterizing local minima of the quadratic program are given. The effect of diagonal perturbations on the number of local minimizers is investigated using a test problem from the literature.

Key words. graph partitioning, min-cut, max-cut, quadratic programming, optimality conditions, graph Laplacian, edge separators, Fiedler vector

AMS subject classifications. 90C35, 90C27, 90C20

PII. S0895480199335829

1. Introduction.

This paper analyzes a continuous quadratic programming formulation for min-cut graph partitioning problems where we partition the vertices of a graph into disjoint sets satisfying specified size constraints, while minimizing the sum of the weights of edges connecting vertices in different sets. As a special case, the discrete quadratic programming formulation of Goemans and Williamson [22] for the max-cut problem is equivalent to a continuous quadratic program in which their discrete variables taking values -1 or $+1$ are replaced by continuous variables with values between -1 and $+1$. Graph partitioning problems arise in circuit board and microchip design, in other layout problems (see [33]), and in sparse matrix pivoting strategies. In parallel computing, graph partitioning problems arise when tasks are partitioned among processors in order to minimize the communication between processors and balance the processor load. For example, an application of graph partitioning to parallel molecular dynamics simulations is given in [44].

Another graph problem with a quadratic programming formulation is the maximum clique problem. In [36] Motzkin and Strauss show that the size of the largest clique in a graph can be obtained by solving a quadratic programming problem, while Gibbons et al. establish in [20] many interesting properties of this formulation.

A general approach for converting a discrete optimization problem to a continuous problem involves a diagonal perturbation. For example, subtracting a sufficiently large multiple of the identity from the quadratic cost matrix in the quadratic assignment problem yields a concave minimization problem whose local minimizers are extreme points of the feasible set, and whose global minimizers are solutions of the original discrete optimization problem (see the book [38, p. 26] by Pardalos and Rosen and the article [4] by Bazaraa and Sherali). One problem with this concave formulation of a discrete minimization problem is that the continuous problem can have many local minimizers. When a continuous optimization algorithm is applied, any of these local minima can trap the iterates. Our approach is related in the sense that we modify

*Received by the editors January 1, 1999; accepted for publication (in revised form) May 28, 1999; published electronically October 19, 1999. This research was supported by the National Science Foundation.

http://www.siam.org/journals/sidma/12-4/33582.html

†Department of Mathematics, University of Florida, Gainesville, FL 32611 (hager@math.ufl.edu).

the diagonal of the cost function. However, we are able to capture the solution of the discrete problem without modifying the cost function to the extent that it becomes concave. By restricting the size of the modification, the number of local minimizers that are candidates for a global minimizer is reduced substantially.

Various approaches to the graph partitioning problem appear in the literature. The seminal paper in this area is that of Kernighan and Lin [31] which presents the problem, application areas, and an exchange algorithm for obtaining approximate solutions. Four classes of algorithms have emerged for the graph partitioning problem:

(a) spectral methods, such as those in [29] and [40], where an eigenvector corresponding to the second smallest eigenvalue (Fiedler vector) of the graph's Laplacian is used to approximate the best partition;
(b) geometric methods, such as those in [21], [28], and [35], where geometric information for the graph is used to find a good partition;
(c) multilevel algorithms, such as those in [13], [14], [30], and [32], that first coarsen the graph, partition the smaller graph, then uncoarsen to obtain a partition for the original graph;
(d) optimization-based methods, such as those in [5], [6], [7], [18], and [45], where approximations to the best partitions are obtained by solving optimization problems.

See [3] for a survey of results in this area prior to 1995.

Here we focus on optimization-based formulations. Much of the earlier work in this area involves relaxations in which constraints are dropped in an optimization problem to obtain a tractable problem whose optimal solution is a lower bound for the optimal partition (see, for example, [6], [17], [41]). We also mention the work of Barnes [5] in which a spectral decomposition of the adjacency matrix is used with the solution of a related transportation problem (linear cost function and linear constraints) to approximate the best partition. In [7] a diagonal perturbation of the adjacency matrix is used to make it positive definite, and a Cholesky factorization of this perturbed matrix leads to a transportation problem whose solution again approximates the best partition. In contrast, our quadratic program is an exact formulation of the original problem in the sense that it has a minimizer corresponding to the best partition. Since the graph partitioning problem is NP-hard, this exact formulation is, in general, a difficult problem to solve.

In [18] Falkner, Rendl, and Wolkowicz present a quadratic optimization problem with both a quadratic constraint and linear equality and inequality constraints that is equivalent to the graph partitioning problem, and they solve (approximately) problems from the literature using the bundle-trust code of Schramm and Zowe. Their constraints are of the form

$$0 \leq x_i \leq 1, \quad \sum_{i=1}^{n} x_i = m, \quad \sum_{i=1}^{n} x_i^2 = m,$$

which force the solution vector to have 0/1 components. In [45] Wolkowicz and Zhao consider another variation of the quadratic constraint, requiring that $x_i^2 = x_i$, to enforce the 0/1 constraint. A semidefinite programming relaxation of the original problem is solved using a primal-dual interior point method. Our quadratic programming formulation does not have a quadratic constraint; the constraints are simply linear equalities and inequalities. We show that the quadratic program has a solution with 0/1 components, and that there is a connection between the Fiedler vector used by Pothen, Simon, and Liou in [40] to compute edge and vertex separators of small size and a solution to our quadratic programming problem. Our proof of the existence of a 0/1 solution is based on the following principle exposed by Tardella in [43]: If a function is minimized over a polyhedron, and if for each face of the polyhedron there exists a direction in the face along which the function is concave (or quasi concave), then there exists a vertex minimizer.

We briefly outline the paper: Section 2 presents the quadratic programming formulation of the two-set graph partitioning problem. In section 3 we give necessary and sufficient optimality conditions for a local minimizer of the quadratic program. These conditions relate the graph structure and the first-order optimality conditions at the given point. In section 4 we examine the effect of diagonal perturbations on the number of local minimizers using a test problem of Donath and Hoffman [17]. The connection between our quadratic program and the second eigenvector of the graph's Laplacian is studied in section 5. In section 6 we conclude with various generalizations of our results to partitions involving more than two sets, to nonsymmetric matrices, and to more general constraints.

2. Two-set partitions. Let G be a graph with n vertices V:

$$V = \{1, 2, \ldots, n\},$$

and let a_{ij} be a weight associated with the edge (i, j). For each i and j, we assume that $a_{ii} = 0$, $a_{ij} = a_{ji}$, and if there is no edge between i and j, then $a_{ij} = 0$. The sign of the weights is not restricted. Given a positive integer $m < n$, we wish to partition the vertices into two disjoint sets, one with m vertices and the other with $n - m$ vertices, while minimizing the sum of the weights associated with edges connecting vertices in different sets. This optimal partition is called a min-cut. We show that for an appropriate choice of the diagonal matrix $\mathbf{D}$, the min-cut can be obtained by solving the following quadratic programming problem:

$$\begin{array}{rl} \text{minimize} & (\mathbf{1} - \mathbf{x})^{\mathsf{T}}(\mathbf{A} + \mathbf{D})\mathbf{x} \\ \text{subject to} & \mathbf{0} \le \mathbf{x} \le \mathbf{1}, \ \ \mathbf{1}^{\mathsf{T}}\mathbf{x} = m. \end{array} \tag{1}$$

More precisely, for an appropriate choice of $\mathbf{D}$, (1) has a solution $\mathbf{y}$ for which each component is either 0 or 1. The two sets V_1 and V_2 in an optimal partition are given by

$$V_1 = \{i : y_i = 1\} \quad \text{and} \quad V_2 = \{i : y_i = 0\}. \tag{2}$$

The following theorem shows how to choose $\mathbf{D}$.

Theorem 2.1. *If $\mathbf{D}$ is chosen so that*

$$d_{ii} + d_{jj} \ge 2a_{ij} \tag{3}$$

for each i and j, then (1) *has a* 0/1 *solution* $\mathbf{y}$ *and the partition given by* (2) *is a min-cut. Moreover, if for each i and j,*

$$d_{ii} + d_{jj} > 2a_{ij}, \tag{4}$$

then every local minimizer of (1) *is a* 0/1 *vector.*

Proof. Given a solution $\mathbf{y}$ to (1), we now construct a piecewise linear path, taking us from $\mathbf{y}$ to a solution $\mathbf{z}$ of (1) whose components are either 0 or 1. Let $\mathcal{F}(\mathbf{y})$ be the inactive (or free) components of the vector $\mathbf{y}$:

$$\mathcal{F}(\mathbf{y}) = \{i : 0 < y_i < 1\}. \tag{5}$$

Let f be the cost function of (1):

$$f(\mathbf{x}) = (\mathbf{1} - \mathbf{x})^{\mathsf{T}}(\mathbf{A} + \mathbf{D})\mathbf{x}. \tag{6}$$

Either $\mathcal{F}(\mathbf{y})$ is empty, and $\mathbf{z} = \mathbf{y}$, or $\mathcal{F}(\mathbf{y})$ has two or more elements since the constraint $\mathbf{1}^{\mathsf{T}}\mathbf{x} = m$ of (1), where m is an integer, cannot be satisfied when $\mathbf{x}$ has a single noninteger component. If $\mathcal{F}(\mathbf{y})$ has two or more elements, we show that there exists another minimizing point $\bar{\mathbf{y}}$ with $\mathcal{F}(\bar{\mathbf{y}})$ strictly contained in $\mathcal{F}(\mathbf{y})$, and $f(\mathbf{x}) = f(\mathbf{y})$ for all $\mathbf{x}$ on the line segment connecting $\mathbf{y}$ and $\bar{\mathbf{y}}$. Utilizing this property in an inductive fashion, we conclude that there exists a piecewise linear path taking us from any given minimizer $\mathbf{y}$ to another minimizer $\mathbf{z}$ with $\mathcal{F}(\mathbf{z}) = \emptyset$ (that is, all the components of $\mathbf{z}$ are either 0 or 1), and $f(\mathbf{x}) = f(\mathbf{y})$ for all $\mathbf{x}$ on this path.

If $\mathcal{F}(\mathbf{y})$ has two or more elements, then choose two elements i and $j \in \mathcal{F}(\mathbf{y})$, and let $\mathbf{v}$ be the vector all of whose entries are zero except that $v_i = 1$ and $v_j = -1$. For ϵ sufficiently small, $\mathbf{x} = \mathbf{y} + \epsilon\mathbf{v}$ is feasible in (1). Expanding f in a Taylor series around $\mathbf{x} = \mathbf{y}$, we have

$$f(\mathbf{y} + \epsilon\mathbf{v}) = f(\mathbf{y}) - \epsilon^2\mathbf{v}^{\mathsf{T}}(\mathbf{A} + \mathbf{D})\mathbf{v}. \tag{7}$$

The $O(\epsilon)$ term in this expansion disappears since $f(\mathbf{y} + \epsilon\mathbf{v})$ achieves a minimum at $\epsilon = 0$, and the first derivative with respect to ϵ vanishes at $\epsilon = 0$. In addition, from the inequality

$$f(\mathbf{y} + \epsilon\mathbf{v}) \geq f(\mathbf{y}) \quad \text{for all } \epsilon \text{ near } 0,$$

we conclude that the quadratic term in (7) is nonnegative, or equivalently,

$$\mathbf{v}^{\mathsf{T}}(\mathbf{A} + \mathbf{D})\mathbf{v} \;=\; d_{ii}v_i^2 + d_{jj}v_j^2 + 2a_{ij}v_iv_j \;=\; d_{ii} + d_{jj} - 2a_{ij} \;\leq\; 0. \tag{8}$$

Since $d_{ii}+d_{jj}-2a_{ij} \geq 0$ by (3), it follows that $d_{ii}+d_{jj}-2a_{ij} = 0$ and $f(\mathbf{y}+\epsilon\mathbf{v}) = f(\mathbf{y})$ for each choice of ϵ. Let $\bar{\epsilon}$ be the largest value of ϵ for which $\mathbf{x} = \mathbf{y} + \epsilon\mathbf{v}$ is feasible in (1). Defining $\bar{\mathbf{y}} = \mathbf{y} + \bar{\epsilon}\mathbf{v}$, $\mathcal{F}(\bar{\mathbf{y}})$ is strictly contained in $\mathcal{F}(\mathbf{y})$ and $\bar{\mathbf{y}}$ achieves the minimum in (1) since $f(\mathbf{y} + \epsilon\mathbf{v}) = f(\mathbf{y})$ for all ϵ. In summary, for any given solution $\mathbf{y}$ to (1), we can find another solution $\bar{\mathbf{y}}$ with $\mathcal{F}(\bar{\mathbf{y}})$ strictly contained in $\mathcal{F}(\mathbf{y})$ and $f(\mathbf{x}) = f(\mathbf{y})$ for all $\mathbf{x}$ on the line segment connecting $\mathbf{y}$ and $\bar{\mathbf{y}}$. This shows that there exists a 0/1 solution $\mathbf{y}$ of (1).

Now, if $\mathbf{y}$ is a 0/1 vector, then $f(\mathbf{y})$ is equal to the sum of the weights of the edges connecting the sets V_1 and V_2 in (2). Conversely, given a partition of V into disjoint sets V_1 and V_2 and defining $z_i = 1$ for each $i \in V_1$ and $z_i = 0$ for each $i \in V_2$, $f(\mathbf{z})$ is the sum of the weights of the edges connecting V_1 and V_2. Combining these two observations, we conclude that the partition associated with $\mathbf{y}$ is a min-cut.

Finally, suppose that (4) holds, $\mathbf{y}$ is a local minimizer for (1), and $\mathbf{y}$ is not a 0/1 vector. As noted above, $\mathcal{F}(\mathbf{y})$ has two or more elements, and the expansion (7) holds where the quadratic term satisfies (8), contradicting (4). We conclude that $\mathcal{F}(\mathbf{y})$ is empty and $\mathbf{y}$ is a 0/1 vector. □

Note that condition (3) is equivalent to requiring that $\mathbf{f}$ in (6) is concave in the direction $\mathbf{v}$, where $\mathbf{v}$ is the vector all of whose entries are zero except that $v_i = 1$ and $v_j = -1$. Hence, concavity is not assumed over the entire space $\mathbf{R}^n$, only along directions corresponding to the edges of the constraint polyhedron. The technique we use in the proof of Theorem 2.1 to convert a noninteger minimizer to an integer minimizer by moving in the direction of the vector $\mathbf{v}$ is also employed by Ageev and

Sviridenko in [1]. Although the first part of Theorem 2.1 asserts the existence of a 0/1 solution to (1), there are instances where (1) has solutions that are not 0/1 vectors. For example, if the off-diagonal elements of $\mathbf{A}$ are all equal to 1 and $\mathbf{D} = \mathbf{I}$, then any feasible point is optimal.

We now consider a slightly more general form of the graph partitioning problem where we still minimize the sum of weights of edges connecting the two sets. However, the size of a set is specified by upper and lower bounds rather than by a fixed number m. Our quadratic programming formulation of this min-cut problem is the following:

$$\begin{array}{rl} \text{minimize} & (\mathbf{1}-\mathbf{x})^{\mathsf{T}}(\mathbf{A}+\mathbf{D})\mathbf{x} \\ \text{subject to} & \mathbf{0} \le \mathbf{x} \le \mathbf{1}, \;\; l \le \mathbf{1}^{\mathsf{T}}\mathbf{x} \le u, \end{array} \tag{9}$$

where l and u are given integers satisfying $0 \le l < u \le n$. The corresponding generalization of Theorem 2.1 involves an additional constraint on the diagonal elements of $\mathbf{D}$.

COROLLARY 2.2. *If* $\mathbf{D}$ *is chosen so that*

$$d_{ii} + d_{jj} \ge 2a_{ij} \quad \textit{and} \quad d_{ii} \ge 0 \tag{10}$$

for each i *and* j*, then* (9) *has a* 0/1 *solution* $\mathbf{y}$ *and the partition given by* (2) *is a min-cut. Moreover, if for each* i *and* j*,*

$$d_{ii} + d_{jj} > 2a_{ij} \quad \textit{and} \quad d_{ii} > 0, \tag{11}$$

then every local minimizer of (9) *is a* 0/1 *vector.*

Proof. Let $\mathbf{y}$ be a solution of (9) and define $m = \mathbf{1}^{\mathsf{T}}\mathbf{y}$. If m is not an integer, then $l < m < u$ since m lies between the integers l and u. For $i \in \mathcal{F}(\mathbf{y})$, the free set defined in (5), let $\mathbf{e}$ be the vector all of whose entries are zero except that $e_i = 1$. The function $f(\mathbf{y} + \epsilon\mathbf{e})$, where f is defined in (6), has a local minimum at $\epsilon = 0$ since $\mathbf{y}$ is the global minimizer of (9) and $\mathbf{y} + \epsilon\mathbf{e}$ is feasible for small perturbations in ϵ. Expanding in a Taylor series around $\epsilon = 0$ gives

$$f(\mathbf{y} + \epsilon\mathbf{e}) = f(\mathbf{y}) - d_{ii}\epsilon^2. \tag{12}$$

Since $d_{ii} \ge 0$ by (10), it follows that $d_{ii} = 0$ or else the local optimality of $\mathbf{y}$ in (9) is violated. Hence,

$$f(\mathbf{y} + \epsilon\mathbf{e}) = f(\mathbf{y}) \tag{13}$$

for each choice of ϵ.

For each $i \in \mathcal{F}(\mathbf{y})$, we increase y_i until either y_i reaches the upper bound 1 or $\mathbf{1}^{\mathsf{T}}\mathbf{y}$ reach the upper bound u. These adjustments in y_i do not change the value of f due to (13), and after these adjustments, $\mathbf{1}^{\mathsf{T}}\mathbf{y}$ must be an integer. Therefore, without loss of generality, we can assume that $m = \mathbf{1}^{\mathsf{T}}\mathbf{y}$ is an integer. Since $\mathbf{y}$ is a solution of (9) and the feasible set of (1) is contained in the feasible set of (9), we conclude that $\mathbf{y}$ is a solution of (1) as well as (9). By Theorem 2.1, (1) has a 0/1 solution which must be a solution of (9).

Now suppose that (11) holds. If $\mathbf{1}^{\mathsf{T}}\mathbf{y}$ is not an integer, then we must have $l < \mathbf{1}^{\mathsf{T}}\mathbf{y} < u$. By (11) $d_{ii} > 0$ for each i. If $i \in \mathcal{F}(\mathbf{y})$, then according to (12) the local optimality of $\mathbf{y}$ is violated. Hence, we conclude that $\mathbf{1}^{\mathsf{T}}\mathbf{y}$ is an integer that we denote by m, and $\mathbf{y}$ is a local minimizer for (1) as well as for (9). By Theorem 2.1, $\mathbf{y}$ is a 0/1 vector. $\square$

Remark 2.1. The proofs of Theorem 2.1 and Corollary 2.2 involve quadratic expansions of the cost function. The linear terms in these expansions all vanish due to the optimality of $\mathbf{y}$. Hence, both of these results are valid if linear terms are added to the cost functions in (1) and (9) since linear terms do not effect the quadratic terms in the expansions.

Now let us consider various applications of Theorem 2.1. If $a_{ij} = 1$ for each edge of the graph G, then $\mathbf{A}$ is simply the graph's adjacency matrix. And if $\mathbf{y}$ is a 0/1 vector, then $f(\mathbf{y})$ is equal to the number of edges connecting the sets V_1 and V_2 in (2) for the partition associated with $\mathbf{y}$. Notice that when $\mathbf{A}$ is the adjacency matrix of the graph, conditions (3) and (10) are satisfied by taking $\mathbf{D} = \mathbf{I}$.

Let $\mathbf{W}$ be an $n \times n$ symmetric matrix whose elements are nonnegative with $w_{ii} = 0$ for each i and consider the choice $\mathbf{A} = -\mathbf{W}$. Since $\mathbf{A} \le \mathbf{0}$, it follows that the conditions (3) and (10) are satisfied by taking $\mathbf{D} = \mathbf{0}$. Hence, for this choice of $\mathbf{A}$ and for $\mathbf{D} = \mathbf{0}$, the quadratic programs (1) or (9) have 0/1 solutions. Since minimizing f is equivalent to maximizing $-f$, the minimization problem (9) is equivalent to the following max-cut problem:

$$\begin{aligned} &\text{maximize} \quad (\mathbf{1} - \mathbf{x})^{\mathsf{T}}\mathbf{W}\mathbf{x} \\ &\text{subject to} \ \ \mathbf{0} \le \mathbf{x} \le \mathbf{1}, \ \ l \le \mathbf{1}^{\mathsf{T}}\mathbf{x} \le u. \end{aligned} \tag{14}$$

If $l = 0$ and $u = n$, then there are no constraints on the size of the sets in the partition. In [22] the following discrete formulation is given for the weighted max-cut problem without constraints on the set size:

$$\begin{aligned} &\text{maximize} \ \ \frac{1}{2}\sum_{i<j} w_{ij}(1 - z_i z_j) \\ &\text{subject to} \ \ z_i \in \{-1, 1\}, \ \ 1 \le i \le n. \end{aligned} \tag{15}$$

The cost function of this discrete quadratic program is equal to $\frac{1}{4}(\mathbf{1}^{\mathsf{T}}\mathbf{W}\mathbf{1} - \mathbf{z}^{\mathsf{T}}\mathbf{W}\mathbf{z})$, and with the substitution $\mathbf{z} = 2\mathbf{x} - \mathbf{1}$, we obtain the equivalent problem

$$\begin{aligned} &\text{maximize} \ \ (\mathbf{1} - \mathbf{x})^{\mathsf{T}}\mathbf{W}\mathbf{x} \\ &\text{subject to} \ \ x_i \in \{0, 1\}, \ \ 1 \le i \le n. \end{aligned} \tag{16}$$

Taking $l = 0$ and $u = n$, Corollary 2.2 implies that (14) has the same maximum as (16). Moreover, there exists a 0/1 solution $\mathbf{y}$ of (14) for which the associated partition (2) maximizes the sum of the weights of the edges connecting V_1 and V_2. As a consequence, if the constraint $z_i \in \{-1, 1\}$ in (15) is changed to $-\mathbf{1} \le \mathbf{z} \le \mathbf{1}$, then the resulting continuous quadratic program has the same maximum value as the discrete program (15). This property for bound-constrained minimization was observed by Rosenberg [42] in the following context: If a polynomial is linear with respect to each of its variables, then its minimum over a box is attained at one of the vertices. Since $w_{ii} = 0$, the function $\mathbf{z}^{\mathsf{T}}\mathbf{W}\mathbf{z}$ is linear in each variable and Rosenberg's result can be applied.

Graph partitioning problems have application to ordering strategies for sparse matrix factorization. In the minimum degree algorithm, we permute two rows and the same two columns of a symmetric positive definite matrix $\mathbf{P}$ in order to obtain as many zeros as possible in the first column. The column and the row that are moved to the first row and column correspond to the positive component of a 0/1 solution of (9) associated with $l = u = 1$, where $a_{ij} = 1$ if $p_{ij} \neq 0$ and $a_{ij} = 0$ otherwise. Likewise, taking $l = u = n/2$, assuming n is even, we obtain a partitioning akin to nested dissection in which all those columns and rows associated with indices in V_1 are permuted to the front of the matrix. Viewed in this graph partitioning context, another ordering emerges. For example, we could take $l = 1$ and u a number slightly larger than 1 to obtain an ordering similar to minimum degree. Or we could take $l < n/2$ and $u > n/2$ to obtain an ordering similar to nested dissection that allows some freedom in the size of the sets in the partition.

3. Necessary and sufficient optimality conditions. In this section, we formulate necessary and sufficient optimality conditions for the quadratic programs of section 2. For a general quadratic program, deciding whether a given point is a local minimizer is NP-hard (see [37], [39]). On the other hand, for the quadratic program associated with the graph partitioning problem, we show in this section that local optimality can be decided quickly. Given any $\mathbf{x}$ that is feasible in (1), let us define the sets

$$\mathcal{U}(\mathbf{x}) = \{i : x_i = 1\} \quad \text{and} \quad \mathcal{L}(\mathbf{x}) = \{i : x_i = 0\}.$$

Given a scalar λ, we define the vector

$$\boldsymbol{\mu}(\mathbf{x}, \lambda) = (\mathbf{A} + \mathbf{D})\mathbf{1} - 2(\mathbf{A} + \mathbf{D})\mathbf{x} + \lambda\mathbf{1}.$$

We also introduce subsets $\mathcal{U}_0$ and $\mathcal{L}_0$ defined by

$$\mathcal{U}_0(\mathbf{x}, \lambda) = \{i \in \mathcal{U}(\mathbf{x}) : \mu_i(\mathbf{x}, \lambda) = 0\} \quad \text{and} \quad \mathcal{L}_0(\mathbf{x}, \lambda) = \{i \in \mathcal{L}(\mathbf{x}) : \mu_i(\mathbf{x}, \lambda) = 0\}.$$

The first-order optimality (Karush–Kuhn–Tucker) conditions associated with a local minimizer $\mathbf{x}$ of (9) can be written in the following way: For some scalar λ,

$$\mathbf{0} \leq \mathbf{x} \leq \mathbf{1}, \quad \mathbf{1}^{\mathsf{T}}\mathbf{x} = m, \quad \text{and} \quad \mathbf{x} \in \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}, \lambda)), \tag{17}$$

where $\mathcal{N}(\boldsymbol{\mu}) = \mathcal{N}_1(\boldsymbol{\mu}) \times \mathcal{N}_2(\boldsymbol{\mu}) \times \cdots \times \mathcal{N}_n(\boldsymbol{\mu})$ is a set-valued map, and

$$\mathcal{N}_i(\boldsymbol{\mu}) = \begin{cases} R & \text{if } \mu_i = 0, \\ \{1\} & \text{if } \mu_i < 0, \\ \{0\} & \text{if } \mu_i > 0. \end{cases}$$

Here R denotes the set of real numbers. The first two conditions in (17) are the constraints in (1), while the last condition is complementary slackness and stationarity of the Lagrangian.

THEOREM 3.1. *Suppose that* (3) *holds and m is a real number with $0 < m < n$. A necessary and sufficient condition for $\mathbf{y}$ to be a local minimizer in* (1) *is that all of the following hold:*

(P1) *For some λ, the first-order conditions are satisfied at $\mathbf{x} = \mathbf{y}$.*

(P2) *For each i and $j \in \mathcal{F}(\mathbf{y})$, where $\mathcal{F}$ is the free index set defined in* (5), *we have $d_{ii} + d_{jj} = 2a_{ij}$.*

(P3) *Consider the three sets $\mathcal{U}_0(\mathbf{y}, \lambda)$, $\mathcal{L}_0(\mathbf{y}, \lambda)$, and $\mathcal{F}(\mathbf{y})$. For each i and j in two different sets, we have $d_{ii} + d_{jj} = 2a_{ij}$.*

The motivation for (P2) and (P3) follows. Those indices in $\mathcal{U}_0(\mathbf{y}, \lambda)$, $\mathcal{L}_0(\mathbf{y}, \lambda)$, and $\mathcal{F}(\mathbf{y})$ correspond to those components of the multiplier $\boldsymbol{\mu}(\mathbf{y}, \lambda)$ that vanish. If the cost function $f(\mathbf{x})$ in (6) is expanded in a Taylor series around $\mathbf{y}$, then the linear terms in the expansion corresponding to zero multiplier components are all zero. If $\mathbf{v}$ is a vector all of whose components are zero except that $v_i = 1$ and $v_j = -1$, where i and j are indices corresponding to multiplier components that vanish, then $\mathbf{v}^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{v} \ge 0$ by (3). Conditions (P2) and (P3) are devised so that $\mathbf{v}^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{v} = 0$ whenever $\mathbf{v}$ is a feasible direction at $\mathbf{y}$ (if $\mathbf{v}^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{v} > 0$, then $\mathbf{y}$ is no longer a local minimizer).

Proof. If $\mathbf{y}$ is a local minimizer in (1), then the first-order conditions (17) hold automatically, while in the proof of Theorem 2.1, we saw that $d_{ii}+d_{jj} = 2a_{ij}$ for each i and $j \in \mathcal{F}(\mathbf{y})$ — see the discussion around (8). For the remainder of the proof, we let $\boldsymbol{\mu}$ and the various sets $\mathcal{L}$, $\mathcal{U}$, $\mathcal{F}$, $\mathcal{L}_0$, and $\mathcal{U}_0$ stand for $\boldsymbol{\mu}(\mathbf{y},\lambda)$, $\mathcal{L}(\mathbf{y})$, $\mathcal{U}(\mathbf{y})$, $\mathcal{F}(\mathbf{y})$, $\mathcal{L}_0(\mathbf{y},\lambda)$, and $\mathcal{U}_0(\mathbf{y},\lambda)$, respectively. We also define complementary sets

$$\mathcal{L}' = \mathcal{L} \setminus \mathcal{L}_0 \quad \text{and} \quad \mathcal{U}' = \mathcal{U} \setminus \mathcal{U}_0.$$

$\mathcal{L}'$ is the set of indices for which $y_i = 0$ and $\mu_i > 0$, while $\mathcal{U}'$ is the set of indices for which $y_i = 1$ and $\mu_i < 0$.

To establish (P3), we expand the cost function in a Taylor series around $\mathbf{y}$. Let L be the Lagrangian defined by

$$L(\mathbf{x}) = f(\mathbf{x}) + \lambda(\mathbf{1}^\mathsf{T}\mathbf{x} - m) - \sum_{i\in\mathcal{L}} \mu_i x_i - \sum_{i\in\mathcal{U}} \mu_i(x_i - 1),$$

where f is the cost function in (6). By the complementary slackness condition in (17) and by the definition of $\boldsymbol{\mu}$, we have $L(\mathbf{y}) = f(\mathbf{y})$ and $\nabla L(\mathbf{y}) = 0$. Expanding the Lagrangian around $\mathbf{y}$, we have

$$L(\mathbf{y}+\mathbf{z}) = L(\mathbf{y}) + \nabla L(\mathbf{y})\mathbf{z} + \frac{1}{2}\mathbf{z}^\mathsf{T}\nabla^2 L(\mathbf{y})\mathbf{z} = f(\mathbf{y}) - \mathbf{z}^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{z}.$$

It follows that

$$\begin{aligned} f(\mathbf{y}+\mathbf{z}) &= L(\mathbf{y}+\mathbf{z}) - \lambda(\mathbf{1}^\mathsf{T}(\mathbf{y}+\mathbf{z}) - m) + \sum_{i\in\mathcal{L}} \mu_i(y_i+z_i) + \sum_{i\in\mathcal{U}} \mu_i(y_i+z_i-1) \\ &= f(\mathbf{y}) - \mathbf{z}^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{z} - \lambda\mathbf{1}^\mathsf{T}\mathbf{z} + \sum_{i\in\mathcal{L}} \mu_i z_i + \sum_{i\in\mathcal{U}} \mu_i z_i. \qquad (18) \end{aligned}$$

Suppose that $i \in \mathcal{U}_0$ and $j \in \mathcal{F}$ and let $\mathbf{v}$ be the vector all of whose entries are zero except that $v_i = -1$ and $v_j = 1$. The vector $\mathbf{x} = \mathbf{y} + \epsilon\mathbf{v}$ satisfies the constraints of (1) for ϵ sufficiently small, and by the definition of $\mathcal{U}_0$, $\mu_i = 0$. By (18), we have

$$f(\mathbf{y}+\epsilon\mathbf{v}) = f(\mathbf{y}) - \epsilon^2(d_{ii} + d_{jj} - 2a_{ij}).$$

Since $\mathbf{y}$ is a local optimizer in (1), we must have $d_{ii} + d_{jj} \le 2a_{ij}$; while by (3), $d_{ii} + d_{jj} \ge 2a_{ij}$. Hence, $d_{ii} + d_{jj} = 2a_{ij}$. A similar argument can be used for all the other possible ways of choosing i and j from different sets $\mathcal{U}_0$, $\mathcal{L}_0$, and $\mathcal{F}$. This completes the proof of (P3).

Now consider the converse. That is, we assume that (P1)–(P3) all hold and we wish to show that $\mathbf{y}$ is a local minimizer in (1). Suppose that $\mathbf{x}$ satisfies the constraints of (1) and define $\mathbf{z} = \mathbf{x} - \mathbf{y}$, so that $\mathbf{x} = \mathbf{y} + \mathbf{z}$. Let $\mathcal{Z}$ denote the set defined by

$$\mathcal{Z} = \mathcal{F} \cup \mathcal{L}_0 \cup \mathcal{U}_0 = \{i : \mu_i = 0\}, \qquad (19)$$

and let $\mathcal{Z}'$ be the complement:

$$\mathcal{Z}' = \mathcal{L}' \cup \mathcal{U}' = \{i : \mu_i \ne 0\}. \qquad (20)$$

$$\begin{array}{cc} & \begin{array}{cccc} \mathcal{F} & \mathcal{L}_0 & \mathcal{U}_0 & \mathcal{Z}' \end{array} \\ \begin{array}{c} \mathcal{F} \\ \mathcal{L}_0 \\ \mathcal{U}_0 \\ \mathcal{Z}' \end{array} & \left[\begin{array}{cccc} = & = & = & ? \\ = & ? & = & ? \\ = & = & ? & ? \\ ? & ? & ? & ? \end{array} \right] \end{array}$$

FIG. 3.1. *Structure of* $\mathbf{A} + \mathbf{D}$.

In the case that $\mathcal{Z}'$ is nonempty, we define the parameter

$$\sigma = \min \{|\mu_i| : i \in \mathcal{Z}'\},$$

which is positive by the definition of $\mathcal{Z}'$. For the remainder of the proof, we assume that $\mathcal{Z}'$ is nonempty, and at the end of the proof, we point out the adjustments that are needed to handle the case where $\mathcal{Z}'$ is empty. Since $\mathbf{x} = \mathbf{y} + \mathbf{z}$ satisfies the constraints in (1), we have $z_i \geq 0$ and $z_j \leq 0$ for all $i \in \mathcal{L}$ and $j \in \mathcal{U}$. Since $\mu_i \geq 0$ and $\mu_j \leq 0$ for all $i \in \mathcal{L}$ and $j \in \mathcal{U}$, it follows that

$$\sum_{i \in \mathcal{L}} \mu_i z_i + \sum_{j \in \mathcal{U}} \mu_j z_j = \sum_{i \in \mathcal{Z}'} \mu_i z_i \geq \sigma \sum_{i \in \mathcal{Z}'} |z_i|. \tag{21}$$

If $\|\cdot\|$ and $\|\cdot\|_{\mathcal{Z}'}$ denote the vector 1-norms defined by

$$\|z\| = \sum_{i=1}^{n} |z_i| \quad \text{and} \quad \|z\|_{\mathcal{Z}'} = \sum_{i \in \mathcal{Z}'} |z_i|,$$

then the relation (21) can be expressed

$$\sum_{i \in \mathcal{L}} \mu_i z_i + \sum_{j \in \mathcal{U}} \mu_j z_j \geq \sigma \|z\|_{\mathcal{Z}'}. \tag{22}$$

Now let us consider the quadratic term in (18). The structure of $\mathbf{A}$ is depicted in Figure 3.1. In this figure, an equal sign means that for the elements in that part of the matrix, we have $d_{ii} + d_{jj} = 2a_{ij}$, while a question mark means that we do not know anything about the elements in that region. The equal sign in the $(\mathcal{F}, \mathcal{F})$ position corresponds to (P2) while the remaining six equal signs correspond to (P3).

We now make a careful study of the quadratic term in (18) which can be expressed

$$-\mathbf{z}^{\mathsf{T}}(\mathbf{A} + \mathbf{D})\mathbf{z} = -\sum_{i,j \in \mathcal{Z}} a_{ij} z_i z_j - \sum_{(i,j) \notin \mathcal{Z} \times \mathcal{Z}} a_{ij} z_i z_j - \sum_{i=1}^{n} d_{ii} z_i^2.$$

For those (i, j) that lie in the part of the matrix in Figure 3.1 corresponding to the equal signs, the relation $a_{ij} = (d_{ii} + d_{jj})/2$ holds. With this substitution, a little algebra reveals that

$$\begin{aligned} -\mathbf{z}^{\mathsf{T}}(\mathbf{A} + \mathbf{D})\mathbf{z} = {} & -d \left(\sum_{i \in \mathcal{Z}} z_i \right) + \frac{1}{2} \sum_{i,j \in \mathcal{L}_0} (d_{ii} + d_{jj} - 2a_{ij}) z_i z_j \\ & + \frac{1}{2} \sum_{i,j \in \mathcal{U}_0} (d_{ii} + d_{jj} - 2a_{ij}) z_i z_j - \sum_{(i,j) \notin \mathcal{Z} \times \mathcal{Z}} a_{ij} z_i z_j - \sum_{i \in \mathcal{Z}'} d_{ii} z_i^2, \end{aligned} \tag{23}$$

where d is defined by

$$d = \sum_{i \in \mathcal{Z}} d_{ii} z_i. \tag{24}$$

Since $\mathbf{x}$ is feasible in (9), we have $z_i \geq 0$ for all $i \in \mathcal{L}_0$ and $z_i \leq 0$ for all $i \in \mathcal{U}_0$. Since $d_{ii} + d_{jj} \geq 2a_{ij}$ by (3), we deduce that

$$\sum_{i,j \in \mathcal{L}_0} (d_{ii} + d_{jj} - 2a_{ij}) z_i z_j + \sum_{i,j \in \mathcal{U}_0} (d_{ii} + d_{jj} - 2a_{ij}) z_i z_j \geq 0. \tag{25}$$

Hence, we have

$$-\mathbf{z}^{\mathsf{T}}(\mathbf{A} + \mathbf{D})\mathbf{z} \geq -d \left(\sum_{i \in \mathcal{Z}} z_i \right) - \sum_{(i,j) \notin \mathcal{Z} \times \mathcal{Z}} a_{ij} z_i z_j - \sum_{i \in \mathcal{Z}'} d_{ii} z_i^2. \tag{26}$$

Combining the lower bounds (22) and (26), we conclude from (18) that

$$f(\mathbf{y} + \mathbf{z}) \geq f(\mathbf{y}) + \sigma \|z\|_{\mathcal{Z}'} - d \left(\sum_{i \in \mathcal{Z}} z_i \right) - \sum_{(i,j) \notin \mathcal{Z} \times \mathcal{Z}} a_{ij} z_i z_j - \sum_{i \in \mathcal{Z}'} d_{ii} z_i^2. \tag{27}$$

Since both $\mathbf{x} = \mathbf{y}$ and $\mathbf{x} = \mathbf{y} + \mathbf{z}$ satisfy the constraint $\mathbf{1}^{\mathsf{T}}\mathbf{x} = m$, it follows that $\mathbf{1}^{\mathsf{T}}\mathbf{z} = 0$, from which we obtain the relation

$$\sum_{i \in \mathcal{Z}} z_i = - \sum_{i \in \mathcal{Z}'} z_i.$$

Taking absolute values gives

$$\left| \sum_{i \in \mathcal{Z}} z_i \right| \leq \|z\|_{\mathcal{Z}'}.$$

Also, observe that

$$|z_i z_j| \leq \|z\| \|z\|_{\mathcal{Z}'} \quad \text{when} \quad (i,j) \notin \mathcal{Z} \times \mathcal{Z}$$

since either $i \in \mathcal{Z}'$ or $j \in \mathcal{Z}'$. Combining these observations with (27) yields

$$f(\mathbf{y} + \mathbf{z}) \geq f(\mathbf{y}) + \|\mathbf{z}\|_{\mathcal{Z}'} \left(\sigma - c\|\mathbf{z}\| \right), \tag{28}$$

where c is a constant that can be bounded in terms of the elements of $\mathbf{A}$ and $\mathbf{D}$. Hence, when $\|\mathbf{z}\|$ is sufficiently small, $f(\mathbf{y} + \mathbf{z}) \geq f(\mathbf{y})$, which implies that $\mathbf{y}$ is a local minimizer of f.

To conclude, we consider the case where $\mathcal{Z}'$ is empty. In this case, all the components of $\boldsymbol{\mu}$ vanish by (19). Hence, the last two terms in the Taylor expansion (18) vanish, while the $\mathbf{1}^{\mathsf{T}}\mathbf{z}$ term vanishes since both $\mathbf{x} = \mathbf{y}$ and $\mathbf{x} = \mathbf{y} + \mathbf{z}$ satisfy the constraint $\mathbf{1}^{\mathsf{T}}\mathbf{x} = m$. For the quadratic term in (18), the first term in the identity (23) vanishes since $\mathbf{1}^{\mathsf{T}}\mathbf{z} = 0$, the next two terms are nonnegative by (25), and the last two terms are not present since the complement of $\mathcal{Z}$ is empty. Combining these observations, $f(\mathbf{y} + \mathbf{z}) \geq f(\mathbf{y})$ whenever $\mathbf{x} = \mathbf{y} + \mathbf{z}$ is feasible in (1). Hence, $\mathbf{y}$ is a global minimizer for (1). This completes the proof. $\square$

Remark 3.1. For quadratic programming problems and a point $\mathbf{y}$ satisfying the first-order conditions (17), a necessary and sufficient condition for $\mathbf{y}$ to be a local minimizer is the copositivity of the quadratic cost matrix over a certain cone (see [12] and [15]). In our context, this copositivity condition is equivalent to the inequality

$$\mathbf{v}^{\mathsf{T}}(\mathbf{A}+\mathbf{D})\mathbf{v} \le 0$$

whenever $\mathbf{v}$ lies in the set

$$\Gamma = \{\mathbf{v} \in \mathbf{R}^n : \mathbf{1}^{\mathsf{T}}\mathbf{v} = 0, v_i \le 0 \text{ if } y_i = 1, v_i \ge 0 \text{ if } y_i = 0, \mathbf{v}^{\mathsf{T}}(\mathbf{A}+\mathbf{D})(\mathbf{1}-2\mathbf{y}) = 0\}.$$

Utilizing the expansion (18), it can be shown that

$$\Gamma = \{\mathbf{v} \in \mathbf{R}^n : \mathbf{1}^{\mathsf{T}}\mathbf{v} = 0, v_i \le 0 \text{ if } i \in \mathcal{U}_0, v_i = 0 \text{ if } i \in \mathcal{U} \setminus \mathcal{U}_0 \text{ or } i \in \mathcal{L} \setminus \mathcal{L}_0,\\ v_i \ge 0 \text{ if } i \in \mathcal{L}_0\}.$$

With further analysis, analogous to that given in the proof of Theorem 3.1, the copositivity condition is equivalent to (P2) and (P3). Other references concerning copositivity and its application to optimality in quadratic programming include [8], [9], [10], [11], [16], [26], and [27].

Remark 3.2. Continuous optimization algorithms typically converge to a point $\mathbf{y}$ that satisfies the first-order conditions (17). Theorem 3.1 provides two conditions (P2) and (P3) that can be checked to determine whether $\mathbf{y}$ is a local minimizer. Moreover, if $\mathbf{y}$ is not a local minimizer, then careful study of the proof of Theorem 3.1 reveals a direction of descent for the quadratic cost function. In particular, suppose that $d_{ii} + d_{jj} > 2a_{ij}$ for indices i and j described in either (P2) or (P3). Let $\mathbf{v}$ be a vector whose entries are all zero except for entries i and j which are chosen so that $v_i = -v_j$ and $|v_i| = 1$. From (18) it follows that

$$f(\mathbf{y}+\epsilon\mathbf{v}) = f(\mathbf{y}) - (d_{ii} + d_{jj} - 2a_{ij})\epsilon^2 \tag{29}$$

since all the terms linear in $\mathbf{z} = \epsilon\mathbf{v}$ vanish. In any of the following cases, we take $v_i = -1$ and $v_j = 1$: (a) $i, j \in \mathcal{F}(\mathbf{y})$ or (b) $i \in \mathcal{U}_0(\mathbf{y}, \lambda)$ and $j \in \mathcal{F}(\mathbf{y})$ or (c) $i \in \mathcal{U}_0(\mathbf{y}, \lambda)$ and $j \in \mathcal{L}_0(\mathbf{y}, \lambda)$. In the case that $i \in \mathcal{L}_0(\mathbf{y}, \lambda)$ and $j \in \mathcal{F}(\mathbf{y})$, we take $v_i = 1$ and $v_j = -1$. Choosing $\mathbf{v}$ in this way, $\mathbf{x} = \mathbf{y} + \epsilon\mathbf{v}$ is feasible in (1) for $\epsilon > 0$ sufficiently small and by (29) the value of the cost function is strictly smaller.

We now examine the case when a local minimizer is strict. If $\mathcal{V} \subset V$ is a collection of vertices from the graph, let $\mathcal{V}_i$ denote the set of edges formed by i and the elements of $\mathcal{V}$:

$$\mathcal{V}_i = \{(i, j) : j \in \mathcal{V}\}.$$

Given a collection of edges $\mathcal{E}$, let $|\mathcal{E}|$ denote the sum of the weights of the edges:

$$|\mathcal{E}| = \sum_{(i,j)\in\mathcal{E}} a_{ij}.$$

COROLLARY 3.2. *A feasible point* $\mathbf{y}$ *for* (1) *is a strict local minimizer if and only if* $\mathcal{F}(\mathbf{y}) = \emptyset$ *and*

$$\min_{i\in\mathcal{L}(\mathbf{y})} |\mathcal{L}_i(\mathbf{y})| - |\mathcal{U}_i(\mathbf{y})| > \max_{j\in\mathcal{U}(\mathbf{y})} |\mathcal{L}_j(\mathbf{y})| - |\mathcal{U}_j(\mathbf{y})|. \tag{30}$$

Proof. Suppose that $\mathbf{y}$ is a strict local minimizer for (1). That is, $f(\mathbf{x}) > f(\mathbf{y})$ when $\mathbf{x}$ is near $\mathbf{y}$ and $\mathbf{x}$ is feasible in (1). If $\mathcal{F}(\mathbf{y})$ is nonempty, then as seen in the proof of Theorem 2.1, $\mathcal{F}(\mathbf{y})$ has at least two elements. By (P2) of Theorem 3.1, $d_{ii} + d_{jj} = 2a_{ij}$ for each i and $j \in \mathcal{F}(\mathbf{y})$. Letting $\mathbf{v}$ be a vector whose elements are all zero except that $v_i = 1$ and $v_j = -1$, the expansion (7) implies that

$$f(\mathbf{y} + \epsilon\mathbf{v}) = f(\mathbf{y}) \tag{31}$$

for all choices of ϵ. Since this violates the assumption that $\mathbf{y}$ is a strict local minimizer, we conclude that $\mathcal{F}(\mathbf{y})$ is empty. By the first-order conditions (17), we have

$$(\mathbf{A1} - 2\mathbf{Ay})_i + \lambda \geq 0 \geq (\mathbf{A1} - 2\mathbf{Ay})_j + \lambda \tag{32}$$

for all $i \in \mathcal{L}(\mathbf{y})$ and $j \in \mathcal{U}(\mathbf{y})$. Since $\mathcal{F}(\mathbf{y})$ is empty, $(\mathbf{A1})_i = |\mathcal{L}_i(\mathbf{y})| + |\mathcal{U}_i(\mathbf{y})|$ and $(\mathbf{Ay})_i = |\mathcal{U}_i(\mathbf{y})|$. Hence, we have

$$(\mathbf{A1} - 2\mathbf{Ay})_i = |\mathcal{L}_i(\mathbf{y})| - |\mathcal{U}_i(\mathbf{y})|, \tag{33}$$

and (32) yields

$$|\mathcal{L}_i(\mathbf{y})| - |\mathcal{U}_i(\mathbf{y})| \geq |\mathcal{L}_j(\mathbf{y})| - |\mathcal{U}_j(\mathbf{y})|$$

for each $i \in \mathcal{L}(\mathbf{y})$ and $j \in \mathcal{U}(\mathbf{y})$. If equality holds, for some $i \in \mathcal{L}(\mathbf{y})$ and $j \in \mathcal{U}(\mathbf{y})$, then equality must hold in (32) as well:

$$(\mathbf{A1} - 2\mathbf{Ay})_i + \lambda = 0 = (\mathbf{A1} - 2\mathbf{Ay})_j + \lambda.$$

This implies that $i \in \mathcal{L}_0(\mathbf{y})$ and $j \in \mathcal{U}_0(\mathbf{y})$. By (P3) of Theorem 3.1, $d_{ii} + d_{jj} = 2a_{ij}$. Choosing $\mathbf{v}$ as we did earlier, $\mathbf{x} = \mathbf{y} + \epsilon\mathbf{v}$ is feasible in (1) for $\epsilon > 0$ sufficiently small, and (31) holds, which violates strict local optimality.

Conversely, suppose that $\mathcal{F}(\mathbf{y}) = \emptyset$ and (30) holds. In this case, we can choose λ such that

$$|\mathcal{L}_i(\mathbf{y})| - |\mathcal{U}_i(\mathbf{y})| + \lambda > 0 > |\mathcal{L}_j(\mathbf{y})| - |\mathcal{U}_j(\mathbf{y})| + \lambda$$

for each $i \in \mathcal{L}(\mathbf{y})$ and $j \in \mathcal{U}(\mathbf{y})$. Utilizing (33) gives

$$(\mathbf{A1} - 2\mathbf{Ay})_i + \lambda > 0 > (\mathbf{A1} - 2\mathbf{Ay})_j + \lambda,$$

for each $i \in \mathcal{L}(\mathbf{y})$ and $j \in \mathcal{U}(\mathbf{y})$. For this choice of λ, the first-order conditions (17) hold and both $\mathcal{L}_0(\mathbf{y}, \lambda)$ and $\mathcal{U}_0(\mathbf{y}, \lambda)$ are empty. Hence, the set $\mathcal{Z}'$ in (20) is simply

$$\mathcal{Z}' = \{1, 2, \ldots, n\}.$$

In this case, the lower bound (28) implies that $\mathbf{y}$ is a strict local minimizer. □

We now consider the quadratic program (9) with inequality constraints. In this case, the first-order KKT conditions are the following: For some λ,

$$\mathbf{0} \leq \mathbf{x} \leq \mathbf{1}, \quad l \leq \mathbf{1}^{\mathsf{T}}\mathbf{x} \leq u, \quad \mathbf{1}^{\mathsf{T}}\mathbf{x} \in \mathcal{M}(\lambda), \quad \text{and} \quad \mathbf{x} \in \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}, \lambda)), \tag{34}$$

where $\mathcal{M}(\lambda)$ is the set-valued map defined by

$$\mathcal{M}(\lambda) = \begin{cases} R & \text{if } \lambda = 0, \\ \{l\} & \text{if } \lambda < 0, \\ \{u\} & \text{if } \lambda > 0. \end{cases}$$

COROLLARY 3.3. *Suppose that* (10) *holds. A necessary and sufficient condition for* $\mathbf{y}$ *to be a local minimizer in* (9) *is that* (P1)–(P3) *hold along with the following additional condition:*

(P4) *In the case that* $\lambda = 0$ *in the first-order condition* (34), $d_{ii} = 0$ *for each* $i \in \mathcal{F}(\mathbf{y}) \cup \mathcal{L}_0 \cup \mathcal{U}_0$ *if* $l < \mathbf{1}^\mathsf{T}\mathbf{y} < u$, $d_{ii} = 0$ *for each* $i \in \mathcal{F}(\mathbf{y}) \cup \mathcal{U}_0$ *if* $\mathbf{1}^\mathsf{T}\mathbf{y} = u$, *and* $d_{ii} = 0$ *for each* $i \in \mathcal{F}(\mathbf{y}) \cup \mathcal{L}_0$ *if* $\mathbf{1}^\mathsf{T}\mathbf{y} = l$.

Proof. We use the notation introduced in the proof of Theorem 3.1. If $\mathbf{y}$ is a local minimizer in (9), then the first-order conditions (34) hold automatically for some scalar λ. Since $\mathbf{y}$ is a local minimizer in (1) with $m = \mathbf{1}^\mathsf{T}\mathbf{y}$, it follows from Theorem 3.1 that (P2) and (P3) hold as well. If $\lambda = 0$ and $\mathbf{e}$ is a vector whose components are all zero except that $e_i = 1$ for some $i \in \mathcal{Z}$, then the expansion (18) yields

$$f(\mathbf{y} + \epsilon\mathbf{e}) = f(\mathbf{y}) - \epsilon^2 d_{ii}. \tag{35}$$

It follows that the local optimality of $\mathbf{y}$ is violated unless $d_{ii} = 0$ in all the cases cited in (P4).

Conversely, let us assume that (P1)–(P4) all hold. We wish to show that $\mathbf{y}$ is a local minimizer in (9). Given a feasible point $\mathbf{x}$ for (9), define $\mathbf{z} = \mathbf{x} - \mathbf{y}$. In Theorem 3.1, $\mathbf{1}^\mathsf{T}\mathbf{z} = 0$ and consequently, the λ term in (18) disappeared. Now this term needs to be included on the right side of (27) to obtain

$$\begin{aligned} f(\mathbf{y} + \mathbf{z}) \geq f(\mathbf{y}) + \sigma\|z\|_{\mathcal{Z}'} - \lambda\mathbf{1}^\mathsf{T}\mathbf{z} \\ -d\left(\sum_{i\in\mathcal{Z}} z_i\right) - \sum_{(i,j)\notin\mathcal{Z}\times\mathcal{Z}} a_{ij}z_iz_j - \sum_{i\in\mathcal{Z}'} d_{ii}z_i^2. \end{aligned} \tag{36}$$

In the proof of Theorem 3.1, we showed that the last two terms in (36) can be bounded by $c\|\mathbf{z}\|\|\mathbf{z}\|_{\mathcal{Z}'}$. Moreover, utilizing the identity

$$\sum_{i\in\mathcal{Z}} z_i = \mathbf{1}^\mathsf{T}\mathbf{z} - \sum_{i\in\mathcal{Z}'} z_i,$$

(36) yields

$$f(\mathbf{y} + \mathbf{z}) \geq f(\mathbf{y}) + (\sigma - c\|\mathbf{z}\|)\|\mathbf{z}\|_{\mathcal{Z}'} - (d + \lambda)\mathbf{1}^\mathsf{T}\mathbf{z}. \tag{37}$$

If $l < \mathbf{1}^\mathsf{T}\mathbf{y} < u$, then $\lambda = 0$ by (34) and $d = 0$ by (P4). It follows from (37) that $\mathbf{y}$ is a local minimizer. If $\mathbf{1}^\mathsf{T}\mathbf{y} = u$, then $\mathbf{1}^\mathsf{T}\mathbf{z} \leq 0$ when $\mathbf{x} = \mathbf{y} + \mathbf{z}$ is feasible in (9). If $\lambda = 0$, then by (P4), we have

$$d = \sum_{i\in\mathcal{L}_0} d_{ii}z_i \geq 0$$

since $z_i \geq 0$ for each $i \in \mathcal{L}$. Again, by (37) and the relation $\mathbf{1}^\mathsf{T}\mathbf{z} \leq 0$, $\mathbf{y}$ is a local minimizer. If $\lambda > 0$, then by choosing $\|\mathbf{z}\|$ small enough that $d + \lambda > 0$, we see from (37) that $\mathbf{y}$ is a local minimizer (since the last term in (37) is nonnegative). The case $\mathbf{1}^\mathsf{T}\mathbf{y} = l$ is treated in an analogous fashion. This completes the proof. ∎

4. An example. Theorem 2.1 and Corollary 2.2 require that the diagonal elements of $\mathbf{D}$ should be chosen large enough to satisfy (3) and (10), respectively. On the other hand, as we now observe, choosing $\mathbf{D}$ too large can lead to a miserable optimization problem. In the case that $\mathbf{D} = s\mathbf{I}$, the quadratic program (1) becomes

$$\begin{aligned} &\text{minimize } (\mathbf{1} - \mathbf{x})^\mathsf{T}(\mathbf{A} + s\mathbf{I})\mathbf{x} \\ &\text{subject to } \mathbf{0} \leq \mathbf{x} \leq \mathbf{1}, \ \ \mathbf{1}^\mathsf{T}\mathbf{x} = m. \end{aligned} \tag{38}$$

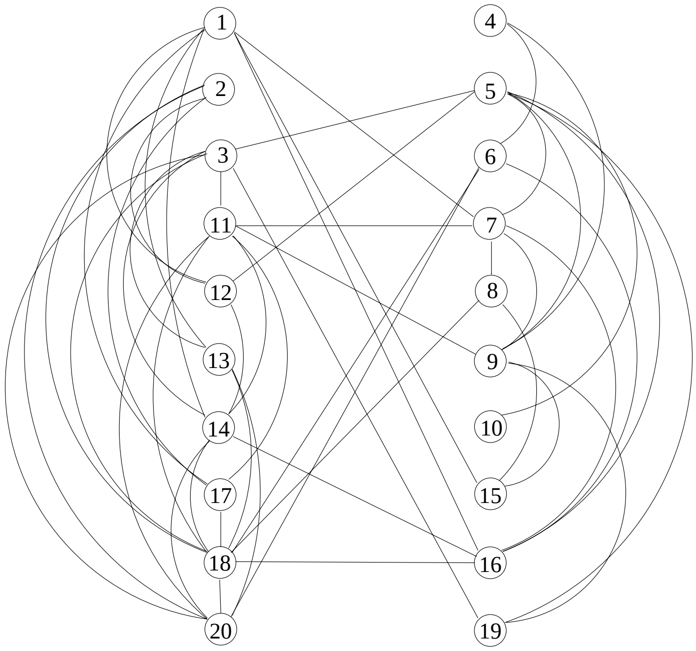

FIG. 4.1. *An example graph.*

Dividing the cost function in (38) by s and taking the limit as s tends to infinity, we obtain the problem

$$\text{(39)} \qquad \begin{array}{c} \text{minimize } (\mathbf{1}-\mathbf{x})^{\mathsf{T}}\mathbf{x} \\ \text{subject to } \ \mathbf{0} \le \mathbf{x} \le \mathbf{1}, \ \ \mathbf{1}^{\mathsf{T}}\mathbf{x} = m, \end{array}$$

The extreme points of the feasible set in either (1) or (38) or (39) is the set

$$\mathcal{X} = \{\mathbf{P}\mathbf{1}_m : \mathbf{P} \in \mathcal{P}\},$$

where $\mathcal{P}$ is the set of $n \times n$ permutation matrices. Since $x = 0$ or $x = 1$ is a strict local minimizer of the function $x(1-x)$, we conclude that any element of $\mathcal{X}$ is a strict local minimizer in the problem (39). In fact, for s sufficiently large, any element of $\mathcal{X}$ is a strict local minimizer in the problem

$$\begin{array}{c} \text{minimize } (\mathbf{1}-\mathbf{x})^{\mathsf{T}}(\mathbf{I} + \frac{1}{s}\mathbf{A})\mathbf{x} \\ \text{subject to } \ \mathbf{0} \le \mathbf{x} \le \mathbf{1}, \ \ \mathbf{1}^{\mathsf{T}}\mathbf{x} = m. \end{array}$$

Hence, as s tends to infinity in (38), every extreme point of the feasible set becomes a local minimizer, and consequently, checking the local minimizers in order to determine the global minimum involves checking every extreme point of the feasible set. As s decreases, fewer of these extreme points become local minimizers in (38), and there are fewer candidates for the global optimum.

As an illustration, let us consider the 20 node graph displayed in Figure 4.1 (see [17, Table 3], [41]) and let $\mathbf{A}$ be the adjacency matrix of the graph. In other words, the weight is 1 for each edge of the graph and 0 otherwise. For this choice of $\mathbf{A}$, we

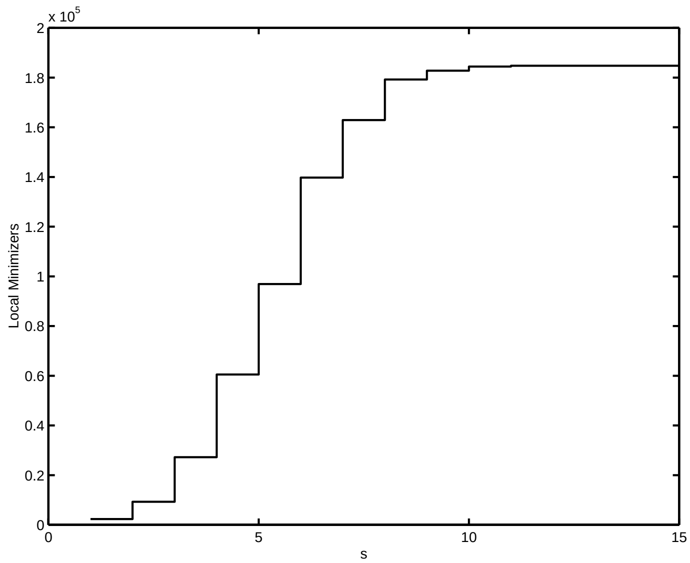

FIG. 4.2. *Number of local minimizers for Figure* 4.1 *graph and optimization problem* (38).

$(\mathbf{1}-\mathbf{x})^{\mathsf{T}}\mathbf{Ax}$	Minimizers	$(\mathbf{1}-\mathbf{x})^{\mathsf{T}}\mathbf{Ax}$	Minimizers
13	2	20	464
14	6	21	440
15	18	22	414
16	36	23	292
17	42	24	164
18	126	25	26
19	304		

FIG. 4.3. *The number of local minimizers in* (1) *for each value of the cost function.*

should choose $s \geq 1$ in (38) to ensure that (3) holds. Following [17], we take $m = 10$, in which case the minimum number of edges separating the two sets of 10 nodes is 13 (the optimal partitioning is shown in Figure 4.1). For the example of Figure 4.1, we computed all the local minimizers of (38) for each value of $s \geq 1$. As s increases, the number of local minimizers increases monotonically. The values of s, where there is a change in the number, are always integers. Figure 4.2 shows the total number of local minimizers as a function of s. The number of local minimizers ranges from 2334 when $1 < s < 2$ up to 184756 for $s \geq 13$. Hence, there are about 79 times as many local minimizers for (38) when $s \geq 13$ as compared to the number of local minimizers when s is between 1 and 2.

For s between 1 and 2, the 2334 local minimizers of (38) yield the distribution of values for the cost function of (1) shown in Figure 4.3. Hence, out of the 2334 local minimizers of (1), only two of them are global minimizers. Note, however, that if we compute any local minimizer of (38), the largest value it can have is 25. Moreover, using 20 iterations of the gradient algorithm (optpack) described in [24], starting from a point near $\mathbf{x} = (m/n)\mathbf{1}$, we converge to a local minimizer of (1) with value $(\mathbf{1}-\mathbf{x})^{\mathsf{T}}\mathbf{Ax} = 14$. Hence, a simple gradient approach provides a partitioning of the vertices that is very close to the optimal partitioning $(\mathbf{1}-\mathbf{x})^{\mathsf{T}}\mathbf{Ax} = 13$.

In contrast, if we take $s = 21$, 11, 6, 3, and 2 in (38) and use exactly the same gradient algorithm and starting point, then we converge to locally minimizing values of 29, 26, 17, 15, and 14, respectively. Thus the smaller values of s yield computed minimizers whose values are closer to the global minimum 13.

The eigenvalues of the matrix $-(\mathbf{A}+\mathbf{I})$ are the following:

$$\begin{array}{rrrrrrr} -7.0429, & -4.1375, & -3.1908, & -2.7637, & -2.4979, & -2.2031, & -1.8808, \\ -1.7844, & -1.3706, & -1.0552, & -0.9066, & -0.4584, & 0.0901, & 0.2508, \\ 0.4315, & 1.0217, & 1.4759, & 1.8608, & 1.9740, & 2.1869 & \end{array}$$

Since there are both positive and negative eigenvalues, the choice $\mathbf{D} = \mathbf{I}$ in the quadratic program (1) has not changed the cost function to the extent that it became concave.

5. Graph eigenvectors. In [40], Pothen, Simon, and Liou propose using an eigenvector associated with the second largest eigenvalue of the Laplacian of a graph in order to compute edge and vertex separators of small size. In this section, we relate this eigenvector to a solution of the quadratic program (1). Let δ_i be the sum of the weights of edges emanating from vertex i:

$$\delta_i = \sum_{j=1}^{n} a_{ij}.$$

(Using the notation of section 3, $\delta_i = |V_i|$.) The Laplacian $\mathbf{L}$ associated with G is defined by

$$l_{ij} = \begin{cases} \delta_i & \text{if } i = j, \\ -a_{ij} & \text{otherwise.} \end{cases}$$

Let $g(\mathbf{x}) = \mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}$ be the quadratic form associated with the Laplacian, and let f be the cost function of the quadratic program (1). See [45] for the first part of the following result.

PROPOSITION 5.1. *We have* $f(\mathbf{x}) = g(\mathbf{x})$ *for all* $\mathbf{x} \in \mathbf{\Omega}$, *where*

$$\mathbf{\Omega} = \{\mathbf{x} \in \mathbf{R}^n : x_i = 0 \text{ or } 1, \quad \mathbf{1}^{\mathsf{T}}\mathbf{x} = m\}.$$

Hence,

$$\min\{f(\mathbf{x}) : \mathbf{0} \le \mathbf{x} \le \mathbf{1},\ \mathbf{1}^{\mathsf{T}}\mathbf{x} = m\} \quad = \quad \min\{g(\mathbf{x}) : \mathbf{x} \in \mathbf{\Omega}\}.$$

Proof. Observe that $(\mathbf{1}-\mathbf{x})^{\mathsf{T}}\mathbf{A}\mathbf{x} = \boldsymbol{\delta}^{\mathsf{T}}\mathbf{x} - \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}$ and

$$\mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x} = \sum_{i=1}^{n} \delta_i x_i^2 - \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}.$$

It follows that

$$f(\mathbf{x}) - g(\mathbf{x}) = \sum_{i=1}^{n} \delta_i (x_i - x_i^2) = 0 \tag{40}$$

for every $\mathbf{x} \in \mathbf{\Omega}$. Since the quadratic program (1) has a solution in $\mathbf{\Omega}$ by Theorem 2.1, the proof is complete. $\square$

By Gerschgorin's theorem (see [23, p. 341] or [25, p. 250]) $\mathbf{L}$ is positive semidefinite and clearly $\mathbf{1}$ is an eigenvector of $\mathbf{L}$ corresponding to the eigenvalue 0. Let $\mathbf{e}_i$, $i = 1, \ldots, n$, denote a linearly independent, normalized set of eigenvectors for $\mathbf{L}$, where $\mathbf{e}_1 = \mathbf{1}/\sqrt{n}$, and where the remaining eigenvectors are ordered so that for the associated eigenvalues, we have

$$0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n.$$

Since $g(\mathbf{1}) = 0$, it follows that for any vector $\mathbf{x}$, $g(\mathbf{x}) = g(\mathbf{Qx})$ where $\mathbf{Q}$ is the projection of a vector onto the orthogonal complement of $\mathbf{1}$. It is easily checked that

$$\mathbf{Qx} = \mathbf{x} - \tfrac{\mathbf{1}^\mathsf{T}\mathbf{x}}{n}\mathbf{1}.$$

Hence, if $\mathbf{x} \in \mathbf{\Omega}$, then

$$\mathbf{Qx} = \mathbf{x} - \tfrac{m}{n}\mathbf{1} \quad \text{and } \|\mathbf{Qx}\|_2 = \sqrt{\tfrac{m(n-m)}{n}}, \tag{41}$$

where $\|\cdot\|_2$ is the Euclidean norm. The function $g(\mathbf{y})$, with $\mathbf{y}$ restricted to a sphere in the orthogonal complement of $\mathbf{1}$, attains its minimum in the eigenspace associated with the second smallest eigenvalue λ_2. By the computation (41), all points of the form $\mathbf{Qx}$ with $\mathbf{x} \in \mathbf{\Omega}$ lie on the sphere of radius $R = \sqrt{m(n-m)/n}$. Hence, the problem of minimizing $g(\mathbf{x})$ over $\mathbf{x} \in \mathbf{\Omega}$ is related to the problem of finding the $\mathbf{x} \in \mathbf{\Omega}$ whose projection onto the orthogonal complement of $\mathbf{1}$ is closest to the eigenspace associated with λ_2.

In [40] the authors focus, in particular, on the case where the vertices are partitioned into two sets of roughly equal size. This case corresponds to taking $m = n/2$ in our notation. Since the eigenvectors associated with the second smallest eigenvalue are all orthogonal to $\mathbf{1}$, the average of the components for any of these eigenvectors is zero. If all the components are of comparable size, then half the components should be positive and the other half should be negative. The $\mathbf{x} \in \mathbf{\Omega}$ for which $\mathbf{Qx}$ is closest to a vector of this form is given by $x_i = 1$ for the positive components and $x_i = 0$ for the negative components. These considerations provide an alternative rationale for the methodology of [40] where the vertices are partitioned according to the sign of the components of an eigenvector corresponding to the second smallest eigenvalue.

This connection, provided by Proposition 5.1 between the quadratics f and g, leads to upper and lower bounds for f over $\mathbf{\Omega}$. In particular, since

$$\lambda_2 \|\mathbf{Qx}\|_2^2 \le g(\mathbf{Qx}) \le \lambda_n \|\mathbf{Qx}\|_2^2,$$

it follows from (41), Proposition 5.1, and the identity $g(\mathbf{x}) = g(\mathbf{Qx})$ for $\mathbf{x} \in \mathbf{\Omega}$ that

$$\lambda_2 R^2 \le \min_{\mathbf{x} \in \mathbf{\Omega}} f(\mathbf{x}) \le \max_{\mathbf{x} \in \mathbf{\Omega}} f(\mathbf{x}) \le \lambda_n R^2, \tag{42}$$

where again $R = \sqrt{m(n-m)/n}$. The following lemma provides a small refinement to these bounds using adjacent eigenvalues:

LEMMA 5.2.

$$\lambda_3 R^2 - (\lambda_3 - \lambda_2)t_2 \le \min_{\mathbf{x} \in \mathbf{\Omega}} f(\mathbf{x}) \le \max_{\mathbf{x} \in \mathbf{\Omega}} f(\mathbf{x}) \le \lambda_{n-1} R^2 + (\lambda_n - \lambda_{n-1})t_n, \tag{43}$$

where

$$t_i = \max_{\mathbf{x} \in \mathbf{\Omega}} \ (\mathbf{e}_i^\mathsf{T}\mathbf{x})^2,$$

with $\mathbf{e}_i$ *the normalized eigenvector associated with the* i*th eigenvalue.*

Proof. We focus on the lower bound since exactly the same procedure can be applied to the upper bound. Given $\mathbf{x} \in \mathbf{\Omega}$, we let z_i denote the coordinates of $\mathbf{Qx}$ relative to the eigenvectors:

$$\mathbf{Qx} = \sum_{i=2}^{n} z_i \mathbf{e}_i.$$

By (41), we have

$$\sum_{i=2}^{n} z_i^2 = R^2 \quad \text{and} \quad z_2^2 = R^2 - \sum_{i=3}^{n} z_i^2.$$

By the definition of g, it follows that

$$\begin{aligned} g(\mathbf{Qx}) &= \sum_{i=2}^{n} \lambda_i z_i^2 \\ &= \lambda_2 R^2 + \sum_{i=3}^{n} (\lambda_i - \lambda_2) z_i^2 \\ &\geq \lambda_2 R^2 + (\lambda_3 - \lambda_2) \sum_{i=3}^{n} z_i^2 \\ &= \lambda_2 R^2 + (\lambda_3 - \lambda_2)(R^2 - z_2^2) \\ &= \lambda_3 R^2 - (\lambda_3 - \lambda_2) z_2^2. \end{aligned}$$

Since $z_2 = \mathbf{e}_2^{\mathsf{T}}\mathbf{x}$, it follows that $z_2^2 \leq t_2$. This completes the proof. ∎

In (42) and (43), we give bounds for the minimum and maximum of f over $\mathbf{\Omega}$ relative to the eigenvalues of the graph Laplacian $\mathbf{L}$. To the extent that the minimum or maximum in (42) or (43) can be evaluated, these inequalities can be used to obtain bounds on the eigenvalues themselves. For example, in the case $m = 1$, the minimum of f over $\mathbf{\Omega}$ is simply the minimum of δ_i, $1 \leq i \leq n$, while the maximum of f over $\mathbf{\Omega}$ is the largest of δ_i, $1 \leq i \leq n$. Letting $\underline{\delta}$ and $\bar{\delta}$ denote the minimum and maximum of the δ_i, we have the estimate (see [19])

$$\lambda_2 \leq \frac{n}{(n-1)}\underline{\delta} \leq \frac{n}{(n-1)}\bar{\delta} \leq \lambda_n.$$

6. Multiset generalizations. In the previous sections, we studied problems that were equivalent to partitioning the vertices of a graph G into two sets of given size, while minimizing the sum of the weights of edges connecting the sets. In this section, we consider the more general problem of partitioning the vertices into k distinct sets $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_k$, with a given number of vertices $m_1, m_2, \ldots, m_k$ in each set, while minimizing the number of edges connecting different sets. Multiset partitions have application in VLSI design (see [3]) and in block iterative techniques for sparse linear systems, where rows and columns are permuted in order to minimize the number of nonzero elements outside the given diagonal blocks.

Let $\mathbf{X}$ be an $n \times k$ matrix, and let us define

$$x_{ij} = \begin{cases} 1 & \text{if } i \in S_j, \\ 0 & \text{if } i \notin S_j. \end{cases}$$

If $\mathbf{x}_j$ is the jth column of $\mathbf{X}$, then the expression $\mathbf{x}_j^\mathsf{T}\mathbf{A}\mathbf{x}_j$ equals the sum of the weights of edges connecting vertices in $\mathcal{S}_j$. The sum of the weights of edges connecting different sets in the partition is minimized when the sum of the weights of edges connecting vertices within the individual sets of the partition is maximized. Hence, the min-cut multiset partitioning problem is equivalent to the following discrete quadratic maximization problem:

$$\begin{array}{ll} \text{maximize} & \text{tr } \mathbf{X}^\mathsf{T}\mathbf{A}\mathbf{X} \\ \text{subject to} & \mathbf{X1} = \mathbf{1}, \quad \mathbf{X}^\mathsf{T}\mathbf{1} = \mathbf{m}, \quad \mathbf{X} \in \mathbf{\Lambda}, \end{array} \tag{44}$$

where tr denotes trace and

$$\mathbf{\Lambda} = \{\mathbf{X} \in \mathbf{R}^{nk} : x_{ij} = 0 \text{ or } 1, \quad 1 \le i \le n, \quad 1 \le j \le k\}.$$

The constraints $\mathbf{X1} = \mathbf{1}$ and $\mathbf{X} \in \mathbf{\Lambda}$ are equivalent to saying that each vertex is contained in precisely one of the sets $\mathcal{S}_j$. The constraint $\mathbf{X}^\mathsf{T}\mathbf{1} = \mathbf{m}$ is equivalent to saying that there are m_j vertices in set $\mathcal{S}_j$ for each j. This discrete quadratic programming formulation of the multiset partitioning problem can be found in [7], for example.

If $\mathbf{X}$ satisfies the constraints of (44), then

$$\text{tr } \mathbf{X}^\mathsf{T}\mathbf{D}\mathbf{X} = \sum_{j=1}^{k} \mathbf{x}_j^\mathsf{T}\mathbf{D}\mathbf{x}_j = \sum_{i=1}^{n} d_i.$$

Consequently, for any choice of the diagonal matrix $\mathbf{D}$, the problem (44) is equivalent to the following problem (since the cost functions differ by a constant, independent of the $\mathbf{X}$):

$$\begin{array}{ll} \text{maximize} & \text{tr } \mathbf{X}^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{X} \\ \text{subject to} & \mathbf{X1} = \mathbf{1}, \quad \mathbf{X}^\mathsf{T}\mathbf{1} = \mathbf{m}, \quad \mathbf{X} \in \mathbf{\Lambda}. \end{array} \tag{45}$$

Our goal in this section is to replace the discrete problem (44), where we impose the constraint $x_{ij} = 0$ or 1, by a continuous problem as in section 2. For example, in the special case $k = 2$, we seek to partition the vertices of the graph into two sets to maximize the total number of edges in the sets. The constraint $\mathbf{X1} = \mathbf{1}$ in (44) implies that $\mathbf{x}_2 = \mathbf{1} - \mathbf{x}_1$, and the cost function in (45) can be rewritten

$$\begin{aligned} \text{tr } \mathbf{X}^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{X} &= \mathbf{x}_1^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{x}_1 + \mathbf{x}_2^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{x}_2 \\ &= -2(\mathbf{1}-\mathbf{x}_1)^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{x}_1 + \mathbf{1}^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{1}. \end{aligned}$$

Hence, after negation and after identifying the $\mathbf{x}$ of (1) with $\mathbf{x}_1$, we see that the cost functions of (45) and of (1) differ only by a constant. Below, the notation $\mathbf{X} \ge \mathbf{0}$ means that every element of the matrix $\mathbf{X}$ is nonnegative.

THEOREM 6.1. *If* $\mathbf{D}$ *is chosen to satisfy* (3)*, then the continuous problem*

$$\begin{array}{ll} \textit{maximize} & \text{tr } \mathbf{X}^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{X} \\ \textit{subject to} & \mathbf{X1} = \mathbf{1}, \quad \mathbf{X}^\mathsf{T}\mathbf{1} = \mathbf{m}, \quad \mathbf{X} \ge \mathbf{0}, \end{array} \tag{46}$$

has a maximizer contained in $\mathbf{\Lambda}$*, and hence, this maximizer is a solution of the discrete problem*

$$\begin{array}{ll} \textit{maximize} & \text{tr } \mathbf{X}^\mathsf{T}(\mathbf{A}+\mathbf{D})\mathbf{X} \\ \textit{subject to} & \mathbf{X1} = \mathbf{1}, \quad \mathbf{X}^\mathsf{T}\mathbf{1} = \mathbf{m}, \quad \mathbf{X} \in \mathbf{\Lambda}. \end{array} \tag{47}$$

Conversely, every solution to (47) *is also a solution to* (46). *Moreover, if* (4) *holds, then every local maximizer for* (46) *lies in* $\mathbf{\Lambda}$.

Proof. Let $\mathbf{Y}$ denote any solution to (46), and let F denote the cost function defined by

$$F(\mathbf{X}) = \text{tr } \mathbf{X}^{\mathsf{T}}(\mathbf{A} + \mathbf{D})\mathbf{X}.$$

If an entry in $\mathbf{Y}$ lies on the open interval $I = (0, 1)$, then we show that there exists another matrix $\bar{\mathbf{Y}}$ with the following properties:

(a) $\bar{\mathbf{Y}}$ is feasible in (46),

(b) $\bar{\mathbf{Y}}$ has at least one fewer entries contained in I than $\mathbf{Y}$, and

(c) $F(\mathbf{X}) = F(\mathbf{Y})$ for all $\mathbf{X}$ on the line segment connecting $\mathbf{Y}$ and $\bar{\mathbf{Y}}$.

Using these properties in an inductive fashion, we obtain, as in the proof of Theorem 2.1, a piecewise linear path taking us from $\mathbf{Y}$ to an optimal point $\mathbf{Z}$ for (46), and the elements of $\mathbf{Z}$ are either 0 or 1.

Proceeding with the construction, if $\mathbf{Y}$ has at least one entry in I, then by interchanging rows and columns if necessary, we can assume, without loss of generality, that $y_{11} \in I$. Since the column sums are integers, there is at least one more entry in column 1 of $\mathbf{Y}$ in I. (No entry of $\mathbf{Y}$ is larger than one since the row sums are all one.) Again, without loss of generality, we assume that $y_{21} \in I$. Since the row sums are integers, there is at least one more entry in the second row $\mathbf{Y}$ in I. Again, without loss of generality, we assume that $y_{22} \in I$.

Continuing this construction, we obtain the piecewise linear path depicted in Figure 6.1, where each point on the path corresponds to an index pair (i, j) for which $y_{ij} \in I$. Eventually, we reach an entry $y_{ij} \in I$ with the property that either the row index i or the column index j agrees with one of the predecessors. As depicted in Figure 6.1, we focus on the case where the row index i agrees with one of the predecessors; an analogous argument applies to the case where the column index agrees with that of a preceding column.

We discard the part of the path in Figure 6.1 that precedes the (i, i) element. Each entry of $\mathbf{Y}$ corresponding to an element of the path lies in I. Let $\mathbf{V}$ be the matrix that is entirely zero except for entries associated with elements on the path. We define $v_{ll} = 1$ for $i \le l \le j$, while the entries of $\mathbf{V}$ corresponding to the other elements on the path are all -1. Since the row and column sums of $\mathbf{V}$ all vanish, $\mathbf{Y} + \epsilon\mathbf{V}$ satisfies the linear constraints of (46) for any choice of ϵ. Since the elements of $\mathbf{Y}$ corresponding to points on the path in Figure 6.1 all lie in I, $\mathbf{Y} + \epsilon\mathbf{V} \ge \mathbf{0}$ for ϵ sufficiently close to 0.

Expanding in a Taylor series, we have

$$F(\mathbf{Y} + \epsilon\mathbf{V}) = F(\mathbf{Y}) + \epsilon^2 F(\mathbf{V}), \tag{48}$$

where the $O(\epsilon)$ term in the expansion vanishes since $F(\mathbf{Y} + \epsilon\mathbf{V})$ attains a local maximum at $\epsilon = 0$. By the structure of $\mathbf{V}$, we have

$$F(\mathbf{V}) = \left(d_{ii} + d_{jj} - 2a_{ij} + \sum_{l=i}^{j-1} d_{ll} + d_{l+1,l+1} - 2a_{l,l+1} \right). \tag{49}$$

By assumption (3), $F(\mathbf{V}) \ge 0$. If $F(\mathbf{V}) > 0$, then the optimality of $\mathbf{Y}$ is contradicted. Hence, $F(\mathbf{V}) = 0$, and we have $F(\mathbf{Y} + \epsilon\mathbf{V}) = F(\mathbf{Y})$ for all choices of ϵ. If $\bar{\epsilon}$ is the first value of ϵ for which a positive component of $\mathbf{Y} + \epsilon\mathbf{V}$ becomes zero, then the matrix

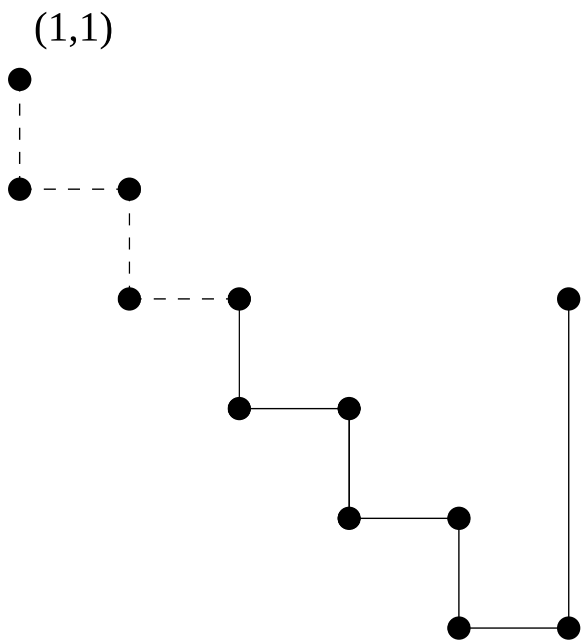

FIG. 6.1. *Indices of entries in* $\mathbf{Y}$ *that lie on the open interval* $(0, 1)$.

$\bar{\mathbf{Y}} = \mathbf{Y} + \bar{\epsilon}\mathbf{V}$ has at least one more zero than $\mathbf{Y}$. This completes the proof of (a)–(c) above.

Now suppose that $d_{ii} + d_{jj} > 2a_{ij}$ for all $i \neq j$ and let $\mathbf{Y}$ be any local maximizer. If $\mathbf{Y}$ has an element in I, then arguing as we did in the first part of the proof, we can construct a matrix $\mathbf{V}$ with elements 0, 1, and -1, whose nonzero elements correspond to elements of $\mathbf{Y}$ in I, and which satisfies (48) and (49). Since the right side of (49) is positive, we contradict the local optimality of $\mathbf{Y}$. Hence, each element of $\mathbf{Y}$ is either 0 or 1. ∎

Theorem 6.1 can be generalized in the following ways:

- Inspecting the proof, we utilize only the fact that the right side of the constraint $\mathbf{X1} = \mathbf{1}$ is an integer; the fact that the integer is 1 is used only to bound the components of $\mathbf{X}$ by 1. Hence, in (46) we can replace the right side of the constraint $\mathbf{X1} = \mathbf{1}$ with a more general vector of positive integers if we add the additional constraint $\mathbf{X} \leq \mathbf{1}$, where, in this matrix setting, $\mathbf{1}$ is the matrix whose elements are all 1.
- The proof of Theorem 6.1 also works if the cost function of (46) is replaced with

$$\sum_{l=1}^{k} \mathbf{x}_l^{\mathsf{T}}(\mathbf{A}_l + \mathbf{D})\mathbf{x}_l,$$

  where $\mathbf{x}_l$ denotes column l of $\mathbf{X}$ and each $\mathbf{A}_l$ is symmetric matrix with zero diagonal that satisfies (3). (In circuit design, we may wish to associate different costs with the edges in different sets.)

- Since the proof of Theorem 6.1 utilizes a Taylor expansion of the quadratic cost function, a linear term can be added to the cost function without changing either the expansions or the conclusions.
- For a nonsymmetric matrix $\mathbf{A}$, we have

$$\text{tr } \mathbf{X}^\mathsf{T}\mathbf{A}\mathbf{X} = \tfrac{1}{2} \text{ tr } \mathbf{X}^\mathsf{T}\mathbf{A}\mathbf{X} + \tfrac{1}{2} \text{ tr } \mathbf{X}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{X} = \text{ tr } \mathbf{X}^\mathsf{T}\mathbf{S}\mathbf{X},$$

where $\mathbf{S} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^\mathsf{T})$ is symmetric. Hence, Theorem 6.1 can be applied to the symmetric matrix $\mathbf{S}$ if the elements satisfy the condition (3). After making the substitution $s_{ij} = (a_{ij} + a_{ji})/2$, we see that the condition (3) of Theorem 6.1 is satisfied if

$$d_{ii} + d_{jj} \geq a_{ij} + a_{ji} \quad \text{for all } i \text{ and } j. \tag{50}$$

Collecting these observations, we have the following corollary.

COROLLARY 6.2. *If $\mathbf{A}_l, l = 1, 2, \ldots, k$, are $n \times n$ matrices, each of which satisfies the condition* (50)*, and $\mathbf{\Phi}$ is a given $k \times n$ matrix, then the continuous problem*

$$\begin{aligned} &\textit{maximize} \quad \text{tr } \mathbf{\Phi}\mathbf{X} + \sum_{l=1}^{k} \mathbf{x}_l^\mathsf{T}(\mathbf{A}_l + \mathbf{D})\mathbf{x}_l \\ &\textit{subject to} \;\; \mathbf{X}\mathbf{1} = \mathbf{r}, \quad \mathbf{X}^\mathsf{T}\mathbf{1} = \mathbf{m}, \quad \mathbf{0} \leq \mathbf{X} \leq \mathbf{1}, \end{aligned} \tag{51}$$

where $\mathbf{r}$ is a vector of positive integers, has a maximizer contained in $\mathbf{\Lambda}$ whenever the feasible set is nonempty, and hence, this maximizer is a solution of the discrete problem

$$\begin{aligned} &\textit{maximize} \quad \text{tr } \mathbf{\Phi}\mathbf{X} + \sum_{l=1}^{k} \mathbf{x}_l^\mathsf{T}(\mathbf{A}_l + \mathbf{D})\mathbf{x}_l \\ &\textit{subject to} \;\; \mathbf{X}\mathbf{1} = \mathbf{r}, \quad \mathbf{X}^\mathsf{T}\mathbf{1} = \mathbf{m}, \quad \mathbf{X} \in \mathbf{\Lambda}. \end{aligned} \tag{52}$$

If $\mathbf{D}$ also satisfies the strict inequality

$$d_{ii} + d_{jj} > a_{ij}^l + a_{ji}^l \quad \textit{for all } i \neq j, \quad 1 \leq l \leq k,$$

where a_{ij}^l is the (i, j)-element of $\mathbf{A}_l$, then every local maximizer of (51) *lies in $\mathbf{\Lambda}$.*

Since the problem (52) with $\mathbf{r} = \mathbf{1}$ and $\mathbf{m} = \mathbf{1}$ is a special case of the quadratic assignment problem, Corollary 6.2 also provides an instance where the quadratic assignment problem can be replaced by a continuous quadratic programming problem whose Hessian is not necessarily positive definite. If $\mathbf{A}_l = \mathbf{0}$ for each l, then (51) is a linear programming problem with *transportation constraints* [34, p. 15]. If $\mathbf{A}_l = \mathbf{0}$ for each l, $\mathbf{X}$ is a square matrix, and $\mathbf{r} = \mathbf{m} = \mathbf{1}$, then (51) is the *linear assignment problem* [2, p. 215]. Hence, for these linear problems, Corollary 6.2 yields, as a special case, the existence of a 0/1 solution. For comparison, the existence of integer solutions in network flow problems can be found, for example, in [2, p. 245].

REFERENCES

[1] A. A. Ageev and M. I. Sviridenko, *Approximation algorithms for maximum coverage and max cut with given sizes of parts*, in Integer Programming and Combinatorial Optimization, Lecture Notes in Comput. Sci. 1610, G. C. Cornuéjols, R. E. Burkard, and G. J. Woeginger, eds., Springer, Berlin, 1999, pp. 17–30.

[2] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network flows*, in Optimization, Handbooks in Oper. Res. and Management Sci. 1, G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, eds., North–Holland, Amsterdam, 1989, pp. 211–369.

[3] C. J. Alpert and A. B. Kahng, *Recent directions in netlist partition: A survey*, Integration, the VLSI Journal, 19 (1995), pp. 1–81.

[4] M. S. Bazaraa and H. D. Sherali, *On the use of exact and heuristic cutting plane methods for the quadratic assignment problem,* J. Oper. Res. Soc., 33 (1982), pp. 991–1003.

[5] E. R. Barnes, *An algorithm for partitioning the nodes of a graph,* SIAM J. Alg. Discrete Methods, 3 (1982), pp. 541–550.

[6] E. R. Barnes and A. J. Hoffman, *Partitioning, spectra, and linear programming,* in Progress in Combinatorial Optimization, W. E. Pulleyblank, ed., Academic Press, New York, 1984, pp. 13–25.

[7] E. R. Barnes, A. Vannelli, and J. Q. Walker, *A new heuristic for partitioning the nodes of a graph,* SIAM J. Discrete Math., 1 (1988), pp. 299–305.

[8] I. M. Bomze, *Block pivoting and shortcut strategies for detecting copositivity*, Linear Algebra Appl., 248 (1996), pp. 161–184.

[9] I. M. Bomze, *Global escape strategies for maximizing quadratic forms over a simplex*, J. Global Optim., 11 (1997), pp. 325–338.

[10] I. M. Bomze and G. Danninger, *A global optimization algorithm for concave quadratic programming problems*, SIAM J. Optim., 3 (1993), pp. 826–842.

[11] I. M. Bomze and G. Danninger, *A finite algorithm for solving general quadratic problems*, J. Global Optim., 4 (1994), pp. 1–16.

[12] J. M. Borwein, *Necessary and sufficient conditions for quadratic minimality*, Numer. Funct. Anal. Optim., 5 (1982), pp. 127–140.

[13] T. N. Bui and C. Jones, *A heuristic for reducing fill-in in sparse matrix factorization,* in Proceedings of the 6th SIAM Conference on Parallel Processing for Scientific Computing, Vol. I, Norfolk, VA, March 22–24, 1993, R. F. Sincovec, D. E. Keyes, M. R. Leuze, L. R. Petzold, and D. A. Reed, eds., SIAM, Philadelphia, 1993, pp. 445–452.

[14] C.-K. Cheng and Y.-C. A. Wei, *An improved two-way partitioning algorithm with stable performance,* IEEE Trans. Comput. Aided Design, 10 (1991), pp. 1502–1511.

[15] L. B. Contesse, *Une caractérisation complète des minima locaux en programmation quadratique*, Numer. Math., 34 (1980), pp. 315–332.

[16] G. Danninger and I. M. Bomze, *Using copositivity for global optimality criteria in concave quadratic programming problems*, Math. Programming, 62 (1993), pp. 575–580.

[17] W. E. Donath and A. J. Hoffman, *Lower bounds for the partitioning of graphs,* IBM J. Res. Develop., 17 (1973), pp. 420–425.

[18] J. Falkner, F. Rendl, and H. Wolkowicz, *A computational study of graph partitioning,* Math. Programming, 66 (1994), pp. 211–240.

[19] M. Fiedler, *Algebraic connectivity of graphs,* Czechoslovak Math. J., 23 (1973), pp. 298–305.

[20] L. E. Gibbons, D. W. Hearn, P. M. Pardalos, and M. V. Ramana, *Continuous characterizations of the maximum clique problem,* Math. Oper. Res., 22 (1997), pp. 754–768.

[21] J. R. Gilbert, G. L. Miller, and S.-H. Teng, *Geometric mesh partitioning: Implementation and experiments,* SIAM J. Sci. Comput., 19 (1998), pp. 2091–2110.

[22] M. X. Goemans and D. P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. ACM, 42 (1995), pp. 1115–1145.

[23] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1989.

[24] W. W. Hager, *Analysis and implementation of a dual algorithm for constrained optimization*, J. Optim Theory Appl., 79 (1993), pp. 427–462.

[25] W. W. Hager, *Applied Numerical Linear Algebra*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[26] J. B. Hiriart-Urruty, *Refinements of necessary optimality conditions in nondifferentiable programming.* I, Appl. Math. Optim., 5 (1979), pp. 63–82.

[27] J. B. Hiriart-Urruty, *Refinements of necessary optimality conditions in nondifferentiable programming.* II. *Optimality and stability in mathematical programming,* Math. Programming Stud., 19 (1982), pp. 120–139.

[28] M. T. Heath and P. Raghavan, *A Cartesian parallel nested dissection algorithm,* SIAM J. Matrix Anal. Appl., 16 (1995), pp. 235–253.
[29] B. Hendrickson and R. Leland, *An improved spectral graph partitioning algorithm for mapping parallel computations,* SIAM J. Sci. Comput., 16 (1995), pp. 452–469.
[30] B. Hendrickson and R. Leland, *A Multilevel Algorithm for Partitioning Graphs,* Technical Report SAND93-1301, Sandia National Laboratories, Albuquerque, NM, 1993.
[31] B. W. Kernighan and S. Lin, *An efficient heuristic procedure for partitioning graphs,* Bell System Tech. J., 49 (1970), pp. 291–307.
[32] G. Karypis and V. Kumar, *A fast and high quality multilevel scheme for partitioning irregular graphs,* SIAM J. Sci. Comput., 20 (1998), pp. 359–392.
[33] T. Lengauer, *Combinatorial Algorithms for Integrated Circuit Layout*, John Wiley, Chichester, 1990.
[34] D. G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA, 1984.
[35] G. L. Miller, S.-H. Teng, W. Thurston, and S. A. Vavasis, *Automatic mesh partitioning,* in Graph Theory and Sparse Matrix Computations, J. Gilbert and J. Liu, eds., IMA Vol. Math. Appl. 56, Springer-Verlag, New York, 1993, pp. 57–84.
[36] T. S. Motzkin and E. G. Strauss, *Maxima for graphs and a new proof of a theorem of Turan,* Canad. J. Math., 17 (1965), pp. 533–540.
[37] K. G. Murty and S. N. Kabadi, *Some NP-complete problems in quadratic and linear programming*, Math. Programming, 39 (1987), pp. 117–129.
[38] P. M. Pardalos and J. B. Rosen, *Constrained Global Optimization: Algorithms and Applications,* Springer-Verlag, Berlin, 1987.
[39] P. M. Pardalos and S. A. Vavasis, *Quadratic programming with one negative eigenvalue is NP-hard*, J. Global Optim., 1 (1991), pp. 15–22.
[40] A. Pothen, H. D. Simon, and K.-P. Liou, *Partitioning sparse matrices with eigenvectors of graphs,* SIAM J. Matrix Anal. Appl., 11 (1990), pp. 430–452.
[41] F. Rendl and H. Wolkowicz, *A projection technique for partitioning the nodes of a graph*, Ann. Oper. Res., 58 (1995), pp. 172–191.
[42] I. Rosenberg, $0-1$ *optimization and nonlinear programming*, Rev. Française Automat. Informat. Recherche Opérationelle, 6 (1972), pp. 95–97.
[43] F. Tardella, *On the equivalence between some discrete and continuous optimization problems*, Ann. Oper. Res., 25 (1990), pp. 291–300.
[44] S.-H. Teng, *Provably good partitioning and load balancing algorithms for parallel adaptive N-body simulation*, SIAM J. Sci. Comput., 19 (1998), pp. 635–656.
[45] H. Wolkowicz and Q. Zhao, *Semidefinite programming relaxations for the graph partitioning problem,* Discrete Appl. Math., to appear.